Sensitive Monogenic Noninvasive Prenatal Diagnosis by Targeted Haplotyping

Vermeulen, Carlo; Geeven, Geert; de Wit, Elzo; Verstegen, Marjon J. A. M.; Jansen, Rumo P. M.; van Kranenburg, Melissa; de Bruijn, Ewart; Pulit, Sara L.; Kruisselbrink, Evelien; Shahsavari, Zahra

Published in:
American Journal of Human Genetics

DOI:
10.1016/j.ajhg.2017.07.012

Publication date:
2017

ARTICLE

Sensitive Monogenic Noninvasive Prenatal Diagnosis by Targeted Haplotyping

Carlo Vermeulen,[1] Geert Geeven,[1] Elzo de Wit,[1,12] Marjon J.A.M. Verstegen,[1] Rumo P.M. Jansen,[2] Melissa van Kranenburg,[1] Ewart de Bruijn,[2] Sara L. Pulit,[2] Evelien Kruisselbrink,[3] Zahra Shahsavari,[4] Davood Omrani,[5] Fatemeh Zeinali,[6] Hossein Najmabadi,[6] Theodora Katsila,[7] Christina Vrettou,[8] George P. Patrinos,[7] Joanne Traeger-Synodinos,[8] Erik Splinter,[9] Jeffrey M. Beekman,[3] Sima Kheradmand Kia,[10] Gerard J. te Meerman,[11] Hans Kristian Ploos van Amstel,[2] and Wouter de Laat[1,*]

During pregnancy, cell-free DNA (cfDNA) in maternal blood encompasses a small percentage of cell-free fetal DNA (cffDNA), an easily accessible source for determination of fetal disease status in risk families through non-invasive procedures. In case of monogenic heritable disease, background maternal cfDNA prohibits direct observation of the maternally inherited allele. Non-invasive prenatal diagnostics (NIPD) of monogenic diseases therefore relies on parental haplotyping and statistical assessment of inherited alleles from cffDNA, techniques currently unavailable for routine clinical practice. Here, we present monogenic NIPD (MG-NIPD), which requires a blood sample from both parents, for targeted locus amplification (TLA)-based phasing of heterozygous variants selectively at a gene of interest. Capture probes-based targeted sequencing of cfDNA from the pregnant mother and a tailored statistical analysis enables predicting fetal gene inheritance. MG-NIPD was validated for 18 pregnancies, focusing on *CFTR*, *CYP21A2*, and *HBB*. In all cases we could predict the inherited alleles with >98% confidence, even at relatively early stages (8 weeks) of pregnancy. This prediction and the accuracy of parental haplotyping was confirmed by sequencing of fetal material obtained by parallel invasive procedures. MG-NIPD is a robust method that requires standard instrumentation and can be implemented in any clinic to provide families carrying a severe monogenic disease with a prenatal diagnostic test based on a simple blood draw.

Introduction

Fragmented DNA expelled by apoptotic cells into the blood plasma is an easily accessible source of biomarkers originating from non-self cells by virtue of their distinct genetic composition. These can be cancer cells carrying a rearranged or mutated genome[1,2] or cells of fetal origin.[3] During pregnancy, a fraction of the maternal cell-free DNA (cfDNA) consists of cell-free fetal DNA.[4,5] This fetal fraction (FF) in maternal cfDNA has enabled non-invasive prenatal testing (NIPT) for aneuploidy to enter routine clinical practice.[6,7] NIPT to detect trisomies is carried out through deep sequencing of maternal cfDNA, followed by a search for significant overrepresentation of fragments originating from a particular chromosome. NIPT circumvents the (perceived) burden of invasive procedures such as chorionic villus sampling (CVS) and amniocentesis, each associated with a small increased risk of miscarriage.[8–10] A similar non-invasive prenatal diagnosis (NIPD) method based on a simple blood draw during pregnancy would be highly beneficial for parents at risk of conceiving a child with a severe monogenic disease. Such a diagnosis would require an accurate assessment of fetal inheritance of point mutations or small indels, but the relatively low contribution of cffDNA to the pool of maternal cfDNA (typically 2%–20%) complicates robust and unambiguous identification of both transmitted alleles through NIPD.[11,12]

Determining the paternally inherited allele in maternal cfDNA is straightforward when the father contributes sequence variants not carried by the mother, as is often the case. Detection of these variants implicitly uncovers the paternally inherited allele.[11,13,14] NIPD for paternally inherited risk alleles is already applied in clinical practice.[11,15,16] Determining which maternal allele is inherited by the fetus is far more challenging, however, since this allele is genetically identical to one of the maternal alleles present in the cfDNA. Identification of the maternally inherited allele therefore requires accurate assessment of which of the two alleles is overrepresented in the cfDNA due to both fetal and maternal contribution versus the allele contributed only maternally. Directly linking the heterozygous SNPs in

[1]Hubrecht Institute-KNAW and University Medical Center Utrecht, Uppsalalaan 8, 3584 CT Utrecht, the Netherlands; [2]Department of Medical Genetics, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, the Netherlands; [3]Department of Pediatric Pulmonology, Wilhelmina Children's Hospital, University Medical Center Utrecht, Lundlaan 6, 3584 EA Utrecht, the Netherlands; [4]Department of Laboratory Medicine, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Arabi Ave, 19839-63113 Tehran, Iran; [5]Department of Medical Genetics, School of Medicine, Shahid Beheshti University of Medical Sciences, Arabi Ave, 1985717443 Tehran, Iran; [6]Kariminejad-Najmabadi Pathology & Genetics Center, #2 Medical Building, Sanat Sq., 14667-13713 Sharak Gharb, Tehran, Iran; [7]Department of Pharmacy, University of Patras University Campus, 26504 Patras-Rio, Greece; [8]Department of Medical Genetics, National and Kapodistrian University of Athens, Choremio Research Laboratory, "Aghia Sophia" Children's Hospital, 11527 Athens, Greece; [9]Cergentis B.V., Yalelaan 62, 3584 CM Utrecht, the Netherlands; [10]Sara Medical Genetics Lab, Shariati St., Niam St., No 53, PO 1948854151 Tehran, Iran; [11]Department of Genetics, University Medical Center Groningen, Hanzeplein 1, 9713 GZ Groningen, the Netherlands
[12]Present address: Division of Gene Regulation, Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, the Netherlands
*Correspondence: w.delaat@hubrecht.eu
http://dx.doi.org/10.1016/j.ajhg.2017.07.012.

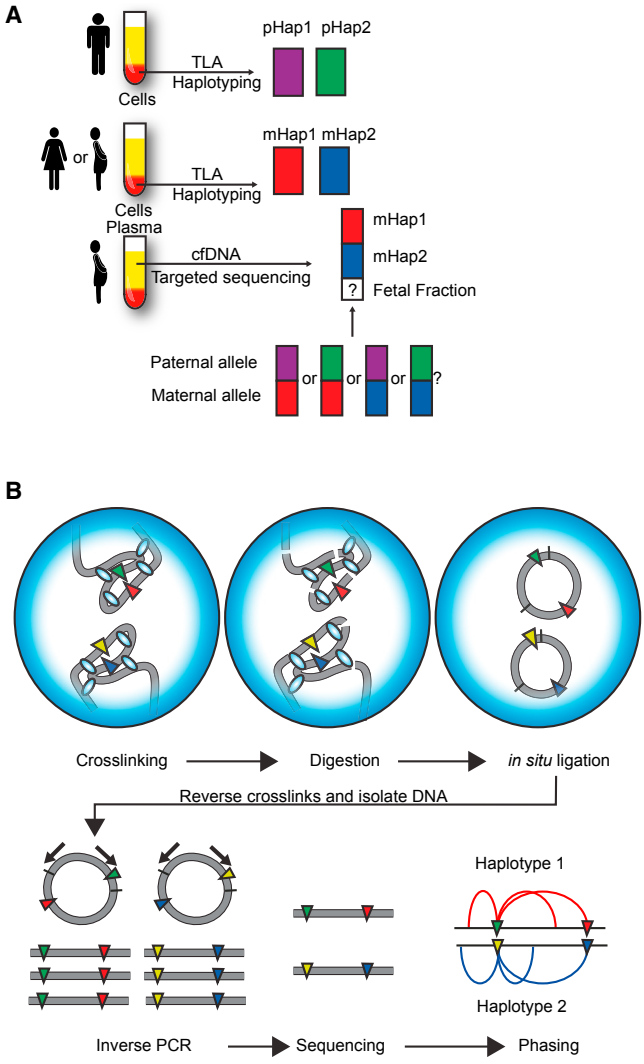

Figure 1. Strategy for Monogenic Non-Invasive Prenatal Diagnosis

(A) Summary of the MG-NIPD approach. Blood is isolated from both parents and cells are used to TLA haplotype the disease locus. Cell-free DNA is isolated from maternal plasma during pregnancy and sequenced to analyze cell-free fetal DNA. Parental locus-specific haplotypes are used to discern which combination of parental alleles is overrepresented in cell-free DNA and therefore inherited by the fetus.

(B) Targeted haplotyping by TLA. Crosslinking (blue ovals), digestion, and proximity ligation primarily yields intra-chromosomal ligation products. Ligation products containing a (viewpoint) SNP of interest (yellow and green triangles) can be selectively amplified by inverse PCR and sequenced. Variants ending up in the same ligation product (indicated by blue and red triangles, respectively) are assigned to the same allele (phasing).

and around the gene of interest to either the mutated or wild-type allele of each parent (called phasing or haplotyping) is thus critical, as the resulting haplotype will yield a high number of informative variants that can be used as a proxy for either the disease or the healthy allele in cfDNA. Identification of many such allele-distinguishing variants, coupled with higher fetal fraction, allows for more robust assessment of the inherited allele in the fetus. Since it is *a priori* unknown how small the fetal fraction will be, a high number of heterozygous variants must be phased to either the disease-linked or wild-type allele. The sequencing and counting of these variants in cfDNA fragments from the pregnant mother then helps determine which of the two alleles is overrepresented in cfDNA and is therefore the maternally inherited allele.[17–20]

For NIPD of monogenic diseases to be routinely applicable in clinical practice, it must be accurate, broadly employable to different monogenic diseases, efficient (in time and resources) in returning results, and cost effective. Strategies that rely on whole-genome haplotyping of the two parents[21–23] are currently too expensive and too analysis intensive to be broadly applicable in the clinic. Furthermore, diagnosis should preferably be feasible without requiring the presence of a proband,[15,16,18,24,25] particularly in a society in which people increasingly have knowledge about their disease carrier status and therefore would want to apply NIPD to their first child. Thus, methods for efficient targeted haplotyping of both parents, combined with targeted deep sequencing of cfDNA, are needed (Figure 1A).

To cost effectively acquire parental haplotypes for NIPD, we implemented the recently developed targeted locus amplification (TLA)[26] method to perform haplotyping specifically around genes of interest. Compared to existing targeted haplotyping strategies, such as long-range PCR or digital (droplet) PCR,[27–29] the TLA technology better enables phasing of dispersed blocks of heterozygous SNPs, even when they are interrupted by long stretches of homozygosity, and is capable of phasing many SNPs per set of primers.[26] cfDNA is then isolated from maternal plasma during pregnancy, enriched for fragments originating from the locus of interest, and sequenced. The haplotype data along with the cfDNA sequence reads are used to determine which haplotypes have been inherited by the fetus. As a proof-of-principle for this approach, we focused on the cystic fibrosis transmembrane regulator (*CFTR* [MIM 602421]) locus, as severe loss-of-function mutations in *CFTR* are known to cause cystic fibrosis (CF [MIM 219700]),[30] and the cytochrome P450 family 21 subfamily A polypeptide 2 (*CYP21A2* [MIM: 613815]) locus, which can contain mutations causal for congenital adrenal hyperplasia (CAH [MIM: 201910]).[31] To investigate the flexibility of our method in additional monogenic diseases, we applied this strategy to ten β-thalassemia (MIM: 613785)-risk families.

Material and Methods

Organoid and Cell Cultures

Leftover rectal biopsies isolated for diagnostic care were used to generate organoid cultures, and informed consent was given for organoid biobanking and the purpose of the study. Organoids were cultured as described.[32] IB3-1 cells were grown as adherent cultures in LHC-8 medium (GIBCO) supplemented with 10% fetal bovine serum and 1% P/S and negatively tested for the presence of mycoplasma.

Sample Preparation

Leftover blood draws were used from anonymous couples for *CFTR* and *CYP21A2* MG-NIPD. The women were at approximately 20 weeks of gestation when fetal anomalies where detected by ultrasound examination. Amniocentesis was performed for diagnostic testing of copy-number variations. High-molecular-weight DNA was isolated from whole blood according to established procedures using a Chemagic Magnetic Separation Module 1 (PerkinElmer). Blood draws were used from β-thalassemia-risk families before CVS procedure. Pregnancies were at approximately 11 weeks gestation (lowest: 7 weeks and 5 days, highest: 11 weeks and 3 days). The use of leftover material for development of new and improved techniques was in accordance with the policy of the UMC Utrecht. In non-Dutch enrollment centers, these studies were approved by the appropriate national ethics authorities.

Additional genomic DNA for SNP genotyping was isolated during TLA template preparation. Plasma was isolated from blood cells by centrifugation of whole blood at 1,600 × *g* and supernatant was subsequently centrifuged at 16,000 × *g* for 10 min and stored at −80°C. cfDNA was then isolated from the blood plasma using a QIAamp DNA Blood Mini Kit (QIAGEN). DNA was isolated from amniotic cells and CVS samples based on phenol extraction according to established procedures.

Viewpoint Genotyping

To identify heterozygous SNPs useable as anchors (viewpoints) to be phased with surrounding variants by TLA, we performed PCR reactions on genomic DNA. FW and SEQ primers (Tables S1 and S2) were combined in PCR reaction using Q5 high fidelity polymerase (New England Biolabs). PCR reactions were pooled per individual and purified using QIAGEN PCR purification columns (QIAGEN). Sanger sequencing was subsequently performed by Macrogen Europe using the SEQ primer.

Targeted Locus Amplification

TLA template was prepared from blood samples as described by De Vree et al.[26] In cases where both cells (for TLA) and cfDNA were isolated from the same blood sample, 5 mL whole blood was centrifuged at 1,600 × *g* for 10 min, and plasma was then used for cfDNA isolation while the cell pellet was resuspended in a 10% fetal bovine serum solution in PBS and treated identical to whole blood for the purpose of template preparation. To create single-cell suspension from organoid cultures, they were treated with trypsin for 5 min at 37°C and disintegrated by pipetting. IB3-1 cells were treated with trypsin until they detached from culture plates and brought into single-cell suspension by pipetting. Single-cell suspensions from organoids or IB3-1 cells were treated identically to isolated white blood cells. PCR was performed as described in De Vree et al.[26] using the FW and RV primers described in Tables S1 and S2, using 100 ng of template per reaction. PCRs were pooled per person prior to tagmentation using nextera XT kit and protocol (Illumina). Tagmented libraries were pooled and sequenced using Illumina miseq, miniseq, or nextseq sequencing platforms with paired-end 150 basepair reads.

Targeted cfDNA Enrichment

A custom Sureselect library was designed (Agilent) for SNPs with >10% heterozygosity in the Dutch population as determined in the Genome of the Netherlands Consortium,[33] the probeset was designed with 5× tiling. For all families, cfDNA pulldown was conducted according to the Sureselect XT2 kit and protocol, with omission of the fragmentation step, since cfDNA is already fragmented. Adaptor mixes were prepared by mixing three stock adapters to maximize unique fragment recovery and diluted 1:20 prior to adaptor ligation, to compensate for the low input quantities. Up to five cfDNAs were pooled in equal amounts after indexing PCR for simultaneous probe hybridization. For the HBB region, SNPs from the dbSNP 1.4.4 database with an average heterozygosity of >10% were selected; probes were designed so that each SNP was covered by six tiled probes, three containing the reference allele, and three containing the alternative allele.

Haplotype Assembly

Raw TLA sequence data were mapped using the BWA SW algorithm to the hg19 human reference genome (UCSC release GRCh37). Heterozygous SNPs were called using a dedicated script, selecting only SNPs with >15% of reads containing the minor allele and a minimum coverage of 25×. Sequence reads containing multiple heterozygous SNP variants were then extracted and each link was counted as described in De Vree et al.[26] Subsequently, a custom haplotyping script was used to construct haplotypes. This script was designed to allow construction of haplotypes, even in the presence of some ambiguous links, where SNPs are found with links to both haplotypes. These links may arise due to sequence errors, PCR artifacts, and perhaps an occasional rare inter-chromosomal contact. Furthermore, the method used here assumes all heterozygous SNPs are bi-allelic, and therefore when one variant is attributed to one haplotype, the other variant can automatically be assigned to the other haplotype (with similar power). In short, the highest covered heterozygous SNP is identified and each variant is assigned to a different haplotype: they serve as fixed starting variants for haplotype assembly (seeds). Subsequently, 25 iterations are performed to stepwise extend the haplotype size, each time adding the most strongly associated (mostly strongly linked and least ambiguous) variants to one of two haplotypes. During the first five iterations, only SNPs where both variants are linked to opposite haplotypes are accepted into the core haplotype, with a strength threshold that decreases per iteration. In the subsequent 15 iterations, only SNPs are accepted where both alleles are found with links to opposite haplotypes without a strength threshold and in the five final iterations SNPs are accepted where both alleles are found with links but with only one allele linking to a haplotype. Linkage of the other allele is assumed in these cases. After the 25th iteration, a final step is added to link poorly covered SNPs, where only one allele is found with links. Here, only SNPs without ambiguous links are accepted and mirrored links are not assumed (Figure S1). We note that allowing ambiguous links increases the size of haplotypes but also increases the chance of erroneous assignments. A small percentage of falsely assigned neutral SNPs does not necessarily affect the predictive power of MG-NIPD but needs to be strictly avoided for disease mutations. Therefore we confirmed that all identified disease mutations as well as the wild-type alleles were indeed both linked unambiguously. Haplotypes shown as "spidergraphs" or clusters show all direct non-ambiguous links used to construct haplotypes. For the analysis of inherited alleles, the two haplotypes are merged and SNPs where only one variant is directly linked but where a reliable (>25× coverage, >15% minor allele frequency [MAF]) heterozygosity call was made are included. For the F508del viewpoint, and the deletions and insertion found in some β-thalassemia carriers, which are not single-nucleotide variants, and therefore are not recognized by the regular pipeline, a modified script was designed to extract all sequence reads containing

either the deletion or the wild-type allele. Discordant positions were identified between the two split datasets and added as links to the set of links identified by the regular pipeline.

cfDNA Sequence Data Processing

cfDNA reads were mapped using BWA MEM to a custom SNP-masked genome (hg19) where the reference sequence of all dbsnp 1.4.2 SNPs in the region of interest were masked to avoid a mapping bias. Duplicates were removed per index using samtools rmdup. Since many cfDNA fragments are shorter than 300 bp, and we used PE150 sequencing, we removed overlapping reads using Genome Analysis Toolkit clip-overlap. Pools of the same sample were merged using samtools after processing. Pileup data were also generated using samtools, SNPs with coverage <20 were excluded. We note that, despite mapping to a masked genome, we still identified a small (~1%) bias toward reference variants in the CFTR analyses. We note that implementing a compensation step for this bias does not change the outcome of any inheritance predictions.

NIPD Analysis

We adapted our NIPD analysis from the RHDO analysis described by Lo et al.[19] Aside from generating haplotypes, a pileup is created from TLA data, for all known dbsnp 1.4.2 SNPs in the region of interest.

For class 1 SNPs, homozygous positions are identified from TLA data with more than 30× coverage and <5% discrepant reads. Of note, this selection may still include heterozygous SNPs in rare cases. From this pileup, class 1 SNPs are selected where opposing alleles are present between parents. For the purpose of FF estimation, we exclude all SNPs where >40% and >20% of sequence reads are from the paternal allele for CFTR and HBB, respectively, since these are most likely maternal non-homozygous SNPs. Class 2 SNPs are identified by comparing the paternal haplotypes with the homozygous maternal SNPs. Observations discrepant to the maternal variant are counted for positions where paternal haplotype 1 would be visible if inherited and for positions where paternal haplotype 2 would be visible if inherited. This typically shows one highly overrepresented allele (Figure 3), both relatively and quantitatively: the paternal haplotype with the highest number of observations is then selected as the inherited paternal haplotype. The inherited paternal haplotype is then extended with all reliably determined homozygous paternal SNPs. Maternal inheritance is based on estimates of the fraction of alleles for the two different types of class 3 maternal SNPs in a given family. We compensate these estimates for overdispersion that is common in high throughput sequence count data, and confidence intervals for fractions of class 3 SNPs are adjusted accordingly. Overdispersion was estimated by comparing the variance of all non-overrepresented SNPs of all families with the theoretical variance of independent observations. We use the corrected estimates of the standard deviation and derived z-scores from the estimated fractions in order to compute clinically relevant posterior risks (Table 1 and Technical Appendix).

Overdispersion

Sequence reads cannot be assumed to be completely independent, due to their PCR-based nature and possibly due to incomplete duplicate removal, resulting in overdispersion. The level of overdispersion in the entire dataset was calculated by comparing the variance of the allele fractions at individual SNPs for which overrepresentation was not expected to the theoretically expected variance in the case of independent observations. The per-family overdispersion was calculated using the variance of a Z transformed proportion, computed for each SNP. When the observed variance of Z is larger than 1.0, this implies overdispersion. We correct for overdispersion in our estimate of the variance of the allele fractions which causes the confidence interval for the proportion of alleles to be wider. The Z score transformed proportion for equally represented alleles is computed over all similarly distributed SNPs using the formula

$$Z = (Fa - 0.5) * 2 * \sqrt{Na + Nb},$$

where Fa is the proportion of alleles A and Na and Nb are the numbers of observed alleles A and B. Let $\widehat{\sigma}_{F\_a}$ denote the sample standard deviation estimate of Fa (assuming independent reads) and let $\widehat{\sigma}_Z$ be the estimate of the SD of Z. Then the overdispersion-corrected SD of Fa becomes $\widehat{\sigma^*_{F\_a}} = \widehat{\sigma}_{F\_a} * \widehat{\sigma}_Z$. We use the overdispersion-corrected SDs in our calculation of the posterior risks.

Posterior Risk Calculation

The posterior risks were calculated using a normal approximation for the distribution of the proportions. The *a posteriori* risk of transmission of a maternal allele to the child uses both type 3A and type 3B SNP data, assuming independence between reads of the two types. If a read contains information on both types of alleles, there is dependency between the two sets. This is a second-order effect that is neglected in the computation because it is rare in practice. Since some of the families in this study do not have a risk allele, the risk of inheritance is calculated for maternal haplotype 2. The likelihoods of the observed data from both SNP types can be computed by multiplication of the appropriate probability densities conditional on the maternal inheritance and for a given the fetal fraction as follows: $\varphi(x \mid \mu,\ \sigma)$ denotes the probability density function of a normally distributed random variable X with mean μ and standard deviation σ. F_A denotes the observed fraction of alleles linked to the maternal risk (or haplotype 2) allele for A SNPs and F_B the same fraction for B SNPs. The overdispersion-corrected standard deviations of these fractions we denote by σ_A and σ_B. The likelihood for type A SNPs, *when the maternal risk allele is transmitted*, is given by $A = \varphi(F_A \mid 0.5, \sigma_A)$ and the likelihood for type B SNPs is given by $B = \varphi(F_B \mid 0.5 + (FF/2),\ \sigma_B)$. When the mother *transmitted the unaffected allele*, the likelihood for type A SNPs is $C = \varphi(F_A \mid 0.5 - (FF/2), \sigma_A)$ and the likelihood of type B SNPs is then $D = \varphi(F_B \mid 0.5,\ \sigma_B)$. These four separate likelihoods can be understood as follows: A measures whether haplotype 2 (or the disease haplotype) is being equally represented in 3A SNPs, and B whether haplotype 2 is being overrepresented at the same time in 3B SNPs. C measures whether maternal haplotype 2 is being underrepresented in 3A SNPs and D whether haplotype 2 is equally represented in 3B SNPs. Since A and B are independently confirming the same hypothesis (mHap2/AF inherited), and C and D are independently confirming the alternate hypothesis (mHap1/WT inherited), we compare A*B and C*D to determine the likelihood of the data under each hypothesis. The posterior risk for a given fetal fraction is thus given by the ratio (AxB) / (AxB+CxD). The fetal fraction is used as weight and integrated out as nuisance parameter, using a density based on the estimated fetal fraction and its overdispersion-corrected variance. For more details and a worked out numerical example, see the Technical Appendix.

Table 1. Overview of MG-NIPD Results

Fam	FF (%)	Class 1	Class 2	Class 3	Tot.	Correct	Unk	Accuracy (%)	Reads	mhap	phap	Post Risk (%)
CFTR 1	26.0	9	190	187	386	378	6	99.4	23,079	2	1	>99.9
CFTR 2	6.0	10	158	229	397	381	14	99.5	19,530	2	2	>99.9
CFTR 3	19	22	62	197	281	275	5	99.6	32,188	1	1	<0.01
CFTR 4	6.1	109	142	163	414	399	13	99.5	28,637	1	1	0.3
CFTR 5	7.6	21	242	86	349	328	15	98.2	9,170	2	1	99.7
CFTR 6	9.4	55	94	138	287	271	12	98.5	16,333	1	1	<0.01
CFTR 7	32.8	9	100	107	216	196	16	98.0	22,823	2	1	>99.9
CFTR 8	17.1	39	105	147	291	267	23	99.6	16,555	2	2	>99.9
CFTR 9	19.7	63	66	197	326	305	17	98.7	35,077	1	2	<0.01
CYP 1	26.0 (*CFTR*)	0	15	183	198	194	1	98.4	19,525	2	1	>99.9
CYP 3	20	19	47	214	280	259	12	96.6	29,423	2	1	>99.9
HBB 1	7.8	11	86	258	355	–	–	–	20,086	AF	AF	99.6
HBB 2	5.8	14	179	300	493	–	–	–	25,770	AF	AF	99.9
HBB 4	9.6	119	304	485	908	762	139	99.0	19,463	WT	WT	<0.01
HBB 5	13.6	1	83	596	680	572	99	98.4	38,099	AF	WT	>99.9
HBB 6	10.5	0	19	515	534	462	55	96.4	44,935	AF	WT	99.8
HBB 7	8.0	6	287	616	909	769	137	99.6	48,448	AF	WT	>99.9
HBB 8	7.2	0	206	241	447	432	9	98.6	21,609	AF	AF	99.3
HBB 9	7.1	142	306	257	705	419	278	98.1	12,460	AF	AF	>99.9
HBB 11	6.9	26	58	165	249	223	16	95.7	20,410	AF	WT	98.1

The outcomes of the MG-NIPD procedure in families 1–9 testing for *CFTR*, *CYP21A2* (in two families, denoted CYP 1 and CYP 3), and the nine included β-thalassemia risk families (*HBB* 1–11, excluding 3 and 10). Listed is the estimated fetal fraction (FF) based on class 1 and 2 SNPs, the number of SNPs identified within each class, and the total number of SNPs for which a genotype was predicted. "Correct" indicates the number of predicted genotypes that were confirmed in the amniocentesis or CVS sample, and "Unk" indicates SNP genotypes that could not be called in the amniocentesis or CVS sample (i.e., were unknown). "Accuracy" indicates the concordance between the genotypes established using TLA and genotypes confirmed through targeted sequencing. "Reads" indicates the raw number of informative sequence reads obtained (for maternal inheritance) from cfDNA in each family. "Mhap" and "Phap" indicate which maternal haplotype (mhap) and paternal haplotype (phap) have been inherited. "Post risk" indicates the posterior risk for inheritance of maternal haplotype 2 (for *CFTR* and *CYP21A2*) or the affected maternal haplotype (for *HBB*).

Theoretical NIPD Requirements

The theoretical required number of reads to reliably detect an overrepresentation of a maternal haplotype for a given FF was calculated based on the expected frequency of overrepresented maternal alleles, while ignoring paternally derived reads.

Fraction inherited hap alleles = Inherited hap allele frequency/total read number

$$Inherited\ haplotype\ allele\ frequency = (1 - FF) + FF$$

$$Non - inherited\ haplotype\ frequency = 1 - FF$$

$$Inherited\ haplotype\ fraction = \frac{(1 - FF) + FF}{(1 - FF) + FF + (1 - FF)} = 1/(2 - FF)$$

The theoretical number of independent and informative reads (N) required to detect a significant deviation from 50% is then calculated:

$$0.5 - 1/(2 - FF) = 3/(2 * \sqrt{(N)})$$

$$N = 9/\left(4 * (0.5 - 1/(2 - \mathrm{FF}))\hat{}2\right)$$

This equation was transformed into the formulas shown in Figure S8.

Graphics

Graphs were made using the R package ggPlot2 (v.1.0.1) and Microsoft Excel. Spiderplots, bar graphs, boxplots, and class 3 distributions were made using R v.3.1.2. Heatmap displays of haplotypes were made using the gplots (v.2.17.0) R package. Clustered haplotype views were made using Cytoscape (v.3.2.1). Sequence data and refseq gene panels were made using Integrated Genome Viewer (v.2.3). Images were further formatted using Adobe illustrator CS6.

Results

We first validated our ability to haplotype the *CFTR* locus by phasing variants in and around the gene. Briefly, TLA uses fixation, digestion, and ligation of DNA fragments that are linearly and physically close together on the chromosome in intact cells. Selective amplification and high throughput sequencing of ligation products formed with

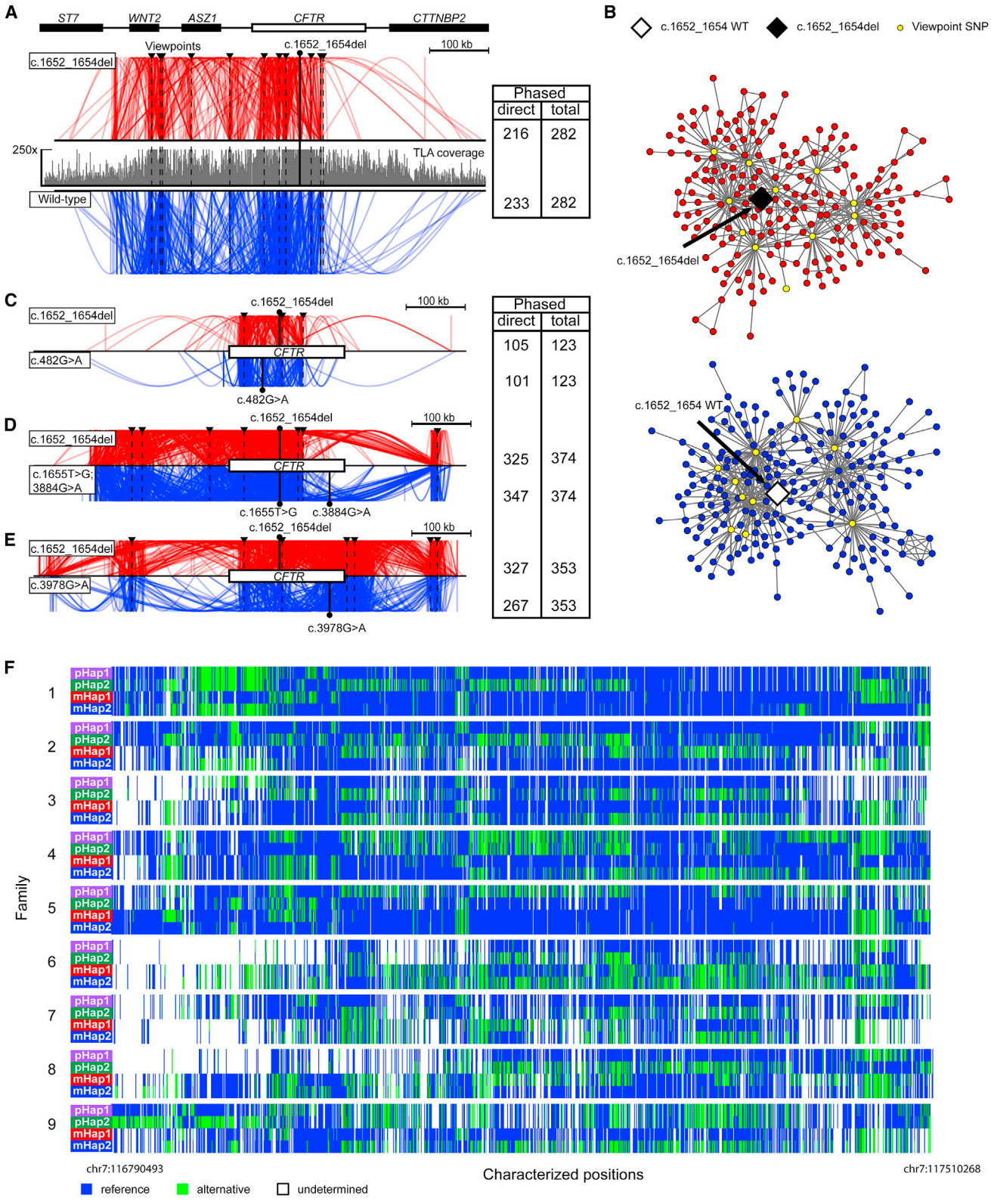

Figure 2. TLA Haplotyping of the *CFTR* Region

(A) TLA haplotyping results obtained from organoids derived from a CF carrier (GenBank: NM_000492.3; c.[1652_1654del];[=]). TLA sequence coverage (plotted in gray, trimmed at 250×) spanned a 710 kb chromosomal interval around the CFTR gene. Informative SNPs in this region inform phasing to construct the c.1652_1654del (top, in red) and unaffected (i.e., wild-type, bottom, in blue) haplotypes. Locations of TLA viewpoints used for this experiment are indicated by black triangles with extended dashed lines. For each allele, the number of independently phased SNPs is shown as well as its total number of phased SNPs (which is based on the merge

(legend continued on next page)

a fragment of interest (called a "viewpoint") allows targeted sequencing of a locus. Since intra-chromosomal crosslinks and ligation events are highly favored over inter-chromosomal events, pairs of SNPs found within the same ligation product can be faithfully assigned to a haplotype (Figure 1B).

We applied TLA to the IB3-1 *CFTR* compound heterozygous cell line (GenBank: NM_000492.3; c.[1652_1654del];[3978G>A])[34] and organoid lines derived from two individuals affected by CF (GenBank: NM_000492.3; c.[1652_1654del];[482G>A] and GenBank: NM_000492.3; c.[1652_1654del];[3884G>A]) and an individual with CF carrier genotype (GenBank: NM_000492.3; c.[1652_1654del];[=]).[32] A series of TLA viewpoints spread across hundreds of kilobases around *CFTR* were designed, each at a common SNP and one at the most recurrent CF variant, the *CFTR*-F508del trinucleotide deletion (GenBank: NM_000492.3; c.[1652_1654del]).[35] PCR and Sanger sequencing were used to verify which of these viewpoint SNPs was heterozygous in a given cell line, a prerequisite for efficient TLA haplotyping, as such SNPs serve as the anchors to be phased to surrounding SNPs by TLA (Table S1 and Figure S2). In all instances, TLA successfully linked the pathogenic variants to many neutral SNPs. Links between individual SNPs spanned up to 600 kb. We identified and assigned 123–374 heterozygous SNPs, spread over approximately 710 kb, to each haplotype (Figures 2A–2E and S3A–S3F). In one of the organoid lines (GenBank: NM_000492.3; c.[1652_1654del];[3884G>A]), an *a priori* unknown additional disease mutation (GenBank: NM_000492.3; c.[1655T>C]) was detected (Figures 2D, S3F, and S3G). TLA further identified common SNPs that were homozygous in a given cell line, which, as explained below, are essential to determine information from cfDNA. We observed a very high similarity between the four F508del haplotypes (Figure S3H) which suggests that the F508del mutation arose in a common ancestor of these persons and provides evidence for the accuracy of TLA haplotyping. We note that the sizes of the identified haplotypes can easily be increased by the design and inclusion of more TLA viewpoints at heterozygous SNPs; such a design was not applied to these test samples. Based on this pilot data, we concluded that our TLA-based strategy for *CFTR* haplotyping could distinguish disease from wild-type *CFTR* alleles and allowed for the phasing of SNPs over hundreds of kilobases of a genomic region of interest.

Subsequently, leftover sample was obtained from nine anonymous couples, where fetal anomalies were detected by ultrasound examination and amniocentesis was performed for diagnostic testing of copy number variations. We used the leftover sample to validate our strategy for non-invasive prenatal *CFTR* diagnostics. Despite there being no pathogenic variants present in these families, determining the combination of haplotypes transmitted to the fetus follows an identical procedure to diagnosing the transmission of a pathogenic variant. A TLA template was prepared from white blood cells. For each individual, the heterozygous SNPs providing informative TLA viewpoints were determined based on PCR and Sanger sequencing (Table S3). TLA haplotyping applied to the 36 *CFTR* alleles of the 18 individuals yielded haplotypes composed of 134–339 SNPs spanning 507–710 kb (Figures 2G, 3A, S4, and S5).

To maximize cfDNA sequencing efficiency, we combined the targeted *CFTR* haplotyping strategy in parents with a targeted sequencing strategy[36] that selectively analyzes the informative *CFTR* sequences present in cfDNA. To this end we designed a capture probe library for the specific pulldown and enrichment of 874 SNPs (cut-off > 10% heterozygosity in the human population of interest[33]) in a 710 kb interval centered on *CFTR*. Barcoded adapters for Illumina sequencing were fused to the purified fragments with a multi-indexing strategy (Figure S6A) and between 5 and 16 million DNA fragments were sequenced from both ends. Fragments with identical start and end position and identical barcodes were considered PCR duplicates and removed from the dataset. Roughly 10% of all sequenced cfDNA fragments were mapped to the region of interest after pulldown, resulting in median coverage (unique informative reads) of 83- to 213-fold across the selected SNPs of the *CFTR* locus (Figures 3B and S6B).

To determine the haplotypes inherited by the fetus, we modified an existing method, called relative haplotype dosage (RHDO) analysis,[19,20] as follows. From parental TLA data, we identified the SNPs at which both parents were homozygous but each for a different allele (class 1 SNPs); at these genomic positions, the father contributes unique variants for which the fetus must be heterozygous (Figure 3C). Such SNPs, if available, allow us to estimate the percentage of cfDNA that was contributed by the fetus, as two times the observed paternally derived fraction. The observed fetal fraction differed widely across the nine

of the two collections of directly phased SNPs, i.e., when allele 1 is linked to haplotype 1, a link is assumed between allele 2 and haplotype 2 and vice versa).

(B) Clustered representation of the constructed c.1652_1654del/WT carrier haplotypes. Each dot represents a SNP allele, shaded with the color representing which of the two haplotypes it has been assigned to (red, disease haplotype; blue, unaffected haplotype). Each line represents one or more ligation products found between two alleles, used to generate the haplotype. Viewpoint SNPs are indicated in yellow.

(C–E) Haplotypes generated for the compound heterozygous CF organoids (C) GenBank: NM_000492.3; c.[1652_1654del];[482G>A] and (D) GenBank: NM_000492.3; c.[1652_1654del];[3884G>A] and the IB3-1 cell line (E) GenBank: NM_000492.3; c.[1652_1654del];[3978G>A].

(F) Schematic representation of the 36 *CFTR* haplotypes constructed in the 9 families included in this study, clustered per family. All haplotypes were expanded to also include identified homozygous SNPs. Blue bars indicate the reference allele (hg19) and green bars represent the alternative variant. White bars represent SNPs where variants were undetermined, due to insufficient TLA coverage.

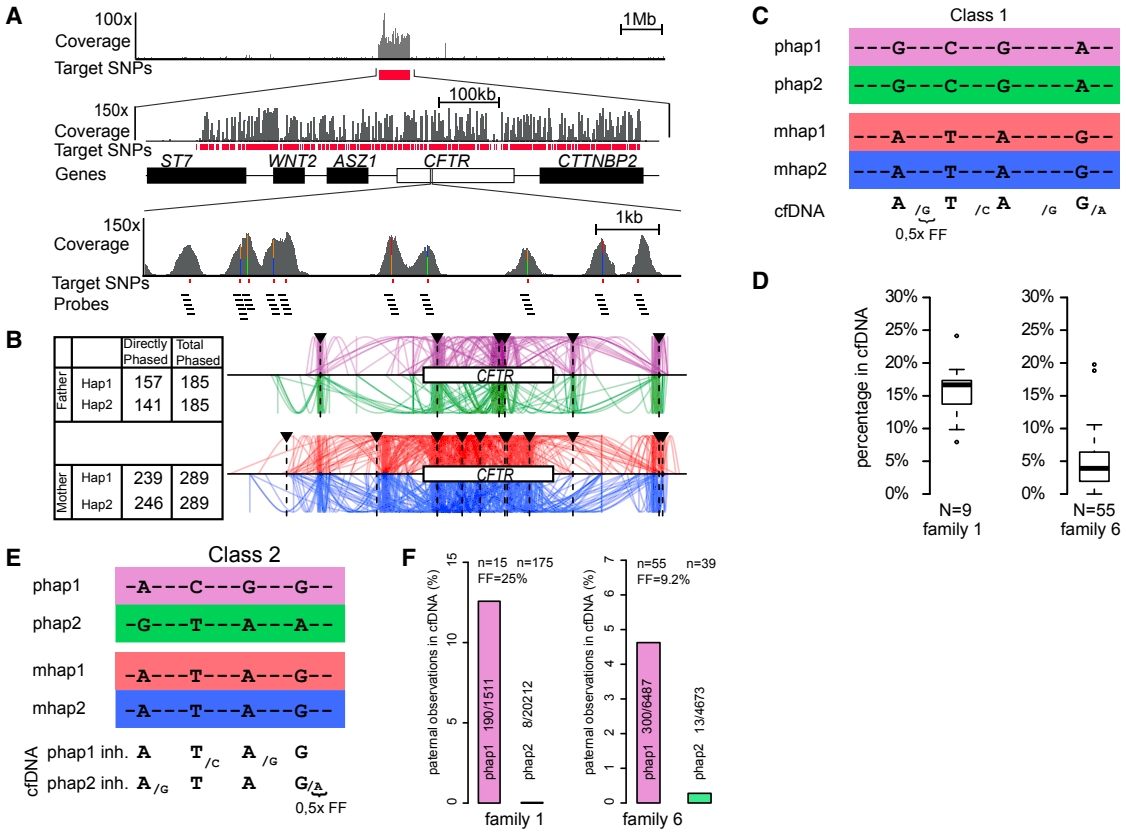

Figure 3. Determining the Fetal Fraction and Paternally Inherited Haplotype
(A) TLA haplotyping results for the parents of family 6.
(B) Sequence coverage obtained by targeted sequencing of cell-free DNA isolated from a pregnant mother (family 6). Across a ~10 Mb chromosomal interval, cell-free DNA sequence coverage almost exclusively localizes to a 710 kb region around *CFTR* (top), where it specifically accumulates at the target SNPs (in red, n = 874) for which hybridization capture probes were designed (middle and bottom).
(C) Class 1 SNPs explained. Parents are homozygous for different variants, which enables estimating the fetal fraction in cell-free DNA.
(D) Boxplots of class 1 SNPs are shown for *CFTR* families 1 and 6. Bold line indicates the median, box indicates first to third quartile, and whiskers extend from the first quartile minus 1.5 times the interquartile range to the third quartile plus 1.5 times the interquartile range.
(E) Class 2 SNPs explained. For these SNPs, the father is heterozygous and the mother is homozygous. Therefore, only one paternal haplotype is detectable in cell-free DNA at each SNP position.
(F) Bar graphs show the class 2 SNPs identified in *CFTR* families 1 and 6 and the percentage of paternal alleles identified for each haplotype. Total read counts are shown inside the bars for each haplotype and the number of individual positions is noted above. The deduced fetal fraction is shown for the inherited paternal haplotype.

pregnant women, ranging from 6.1% to 32.8% (Figures 3D and S9A).

Subsequently, the paternally inherited haplotype is determined using SNPs where the father is heterozygous and the mother is homozygous (class 2 SNPs) (Figure 3E). At each of these positions, one paternal allele is discernible and presence or absence of this variant over all class 2 SNPs reveals which of the two paternal haplotypes was transmitted (Figure 3F). In all nine instances, the paternally inherited *CFTR* allele was readily discernible (Figure S9B). Class 2 SNPs where the paternal haplotype is observed are combined with the class 1 SNPs to more accurately estimate the fetal fraction.

With the fetal fraction, paternally inherited haplotype, and two maternally transmittable haplotypes known, we could then deduce which allele was transmitted by the mother. To identify maternally transmitted alleles, we sub-classified all maternal heterozygous SNPs (class 3 SNPs) depending on whether the variant identical to the variant inherited from the father was carried on maternal haplotype 1 (denoted "type A") or on maternal haplotype 2 (denoted "type B;" Figure 4A). For clarity: this implies that the type A SNP alleles on maternal haplotype 2 and the type B SNP alleles on maternal haplotype 1 are different from those inherited from the father (Figures 4A and 4B). We then calculate for maternal haplotype 2 (the "disease" haplotype when the mother is a carrier, see below) the expected cfDNA ratios of its type A and type B SNPs, first under the assumption that the mother transmits her haplotype 1 and then under the assumption that she transmits her haplotype 2 (Figure 4B). Under the first assumption (maternal haplotype 1 is inherited), haplotype 2 will be under-represented in its type A SNPs while its type B SNPs will be neither enriched nor depleted. Under the second assumption (maternal haplotype 2 is inherited), type A SNPs originating from this haplotype will be neither enriched nor

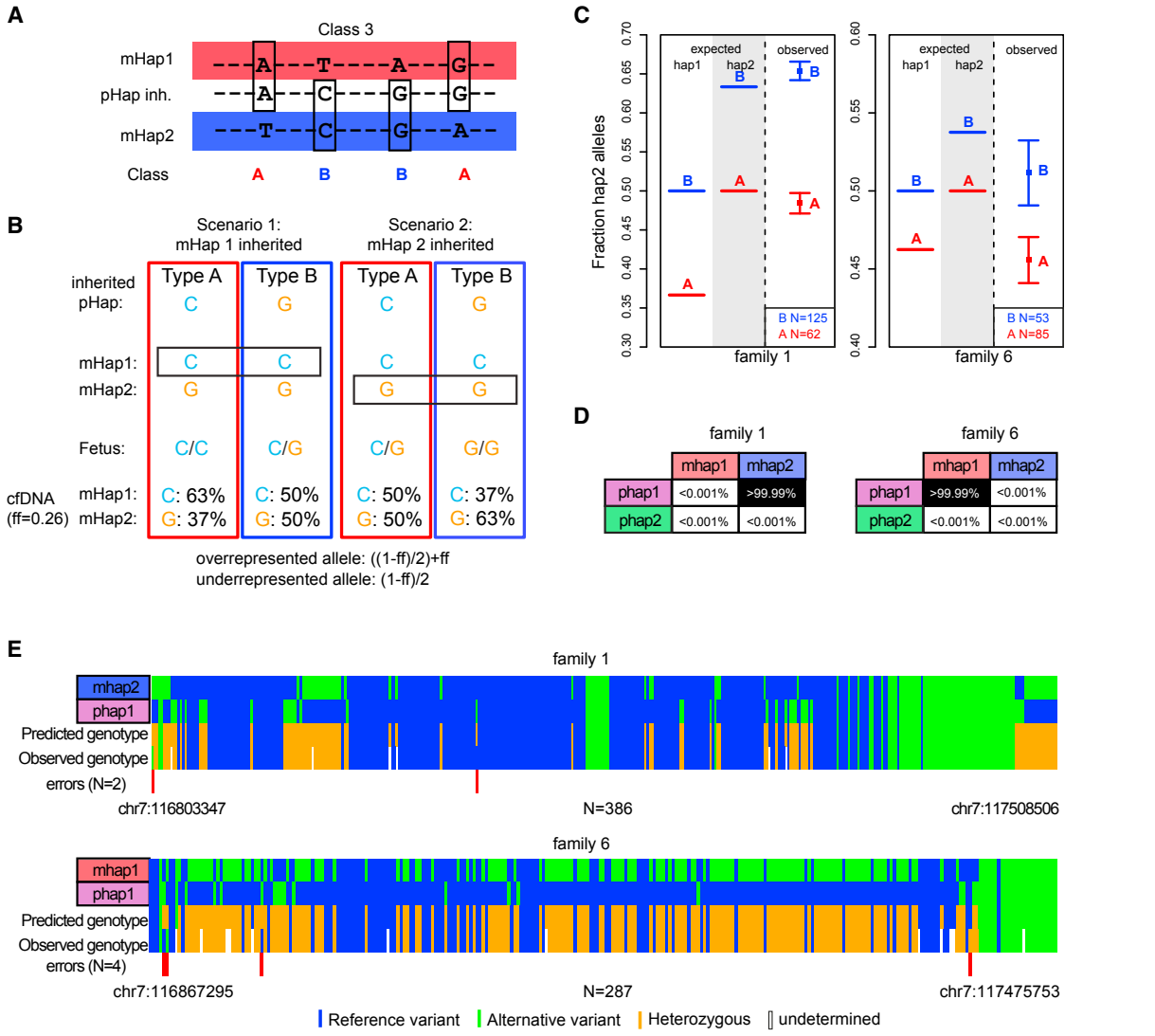

Figure 4. Determining the Maternally Inherited Haplotype

(A) Class 3 SNPs explained. After the paternal inherited haplotype is determined, the maternal heterozygous SNPs are sub-classified in type A SNPs where the maternal haplotype 1 allele is equal to the paternal allele and type B SNPs where the maternal haplotype 2 allele is equal to the paternal allele.

(B) Example calculations for the haplotype representations in type A and B SNPs for a family with a FF of 0.26 (as is the case with family 1). In scenario 1, inheritance of maternal haplotype 1 is assumed, and in scenario 2, inheritance of maternal haplotype 2 is assumed.

(C) The expected distributions of A and B SNPs is made for inheritance of either maternal haplotype 1 or 2 (left of dashed line). The observed distribution (right of dashed line) is then compared to these expected distributions. Plots are shown for *CFTR* family 1 and 6. The number of individual A and B SNPs are indicated in the lower right corner. Squares indicate the mean fraction of hap2 alleles, whiskers indicate the 95% confidence interval, corrected for overdispersion.

(D) Statistical analysis of the observed haplotype 2 fractions in type A and B SNPs yields a probability for each of the four possible combinations of inherited haplotypes. Probabilities are shown for *CFTR* families 1 and 6.

(E) By combining the two expected inherited haplotypes, the fetal genotype was predicted for all involved SNP positions. These predicted genotypes were subsequently compared to the genotypes observed in the amniocentesis sample directly obtained from the fetus, to determine the accuracy of the identified haplotypes, and the validity of the inheritance prediction. Data are shown for families 1 and 6.

depleted, but cfDNA will be enriched for haplotype 2 in type B SNPs (Figures 4B and S7). The expected level of over- or underrepresentation differs per pregnancy and equals half the previously estimated fetal fraction. We subsequently compared the two expected distributions to the observed fraction of haplotype 2 type A and B SNPs, while accounting for overdispersion of allele counts (Figures 4C and S9C). (Due to technical reasons this is common to next-generation sequencing data, but is often ignored; the observed readcounts show a larger variance than would be expected for the statistical models used for analysis. If not properly accounted for, this overdispersion could result in over-estimated confidence levels. See Material and Methods and Technical Appendix.) Statistical testing of the observed versus the expected distributions with the given fetal fraction allowed us to determine the probability of each

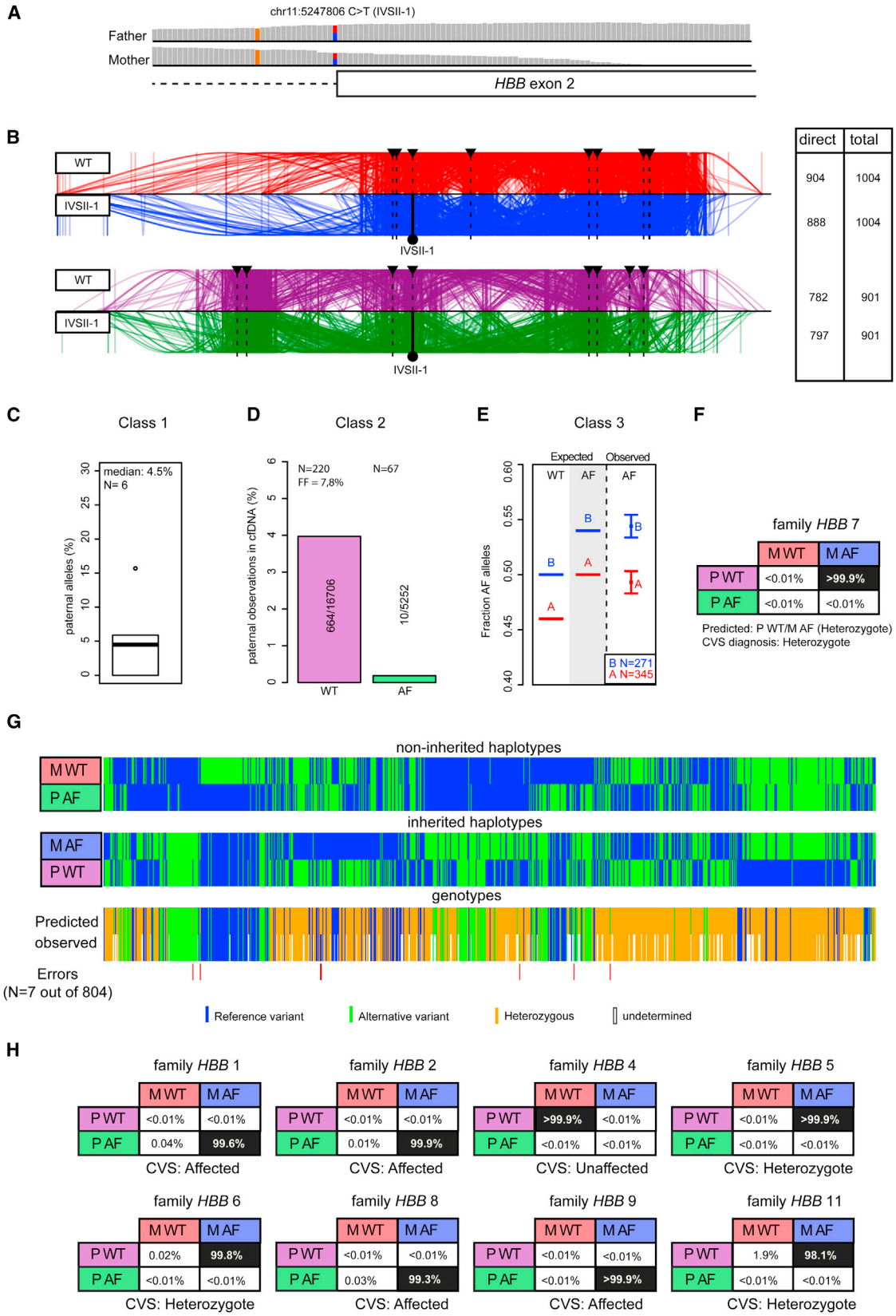

Figure 5. MG-NIPD in a β-Thalassemia Risk Family

A family (*HBB* 7) with an identical mutation at chr11:5347806 (hg19) was admitted for a CVS at 11 weeks of pregnancy. (A) TLA sequencing confirmed the presence of the heterozygous mutation (IVSII-1) at the downstream splice junction of *HBB* exon 2. (B) Haplotypes were constructed in the ~850 kb region surrounding *HBB*, using several viewpoints on heterozygous SNPs and a viewpoint on the HBB gene.

(legend continued on next page)

maternal haplotype being inherited (Figure 4D). The confidence levels (adjusted for overdispersion) were used in a posterior risk calculation for transmission of haplotype 2 (which may be hypothetically considered the disease allele here, but know that these families are not at risk for CF). Using our approach, in all cases we were able to predict allele inheritance in the fetus with >99% confidence (Table 1).

By combining the predicted parental haplotypes, we constructed per fetus a predicted genotype across hundreds of SNPs at the locus of interest. To validate the correctness of our predictions and to assess the accuracy of the parental haplotypes generated by TLA technology, we performed targeted sequencing of the locus of interest using fetal material obtained through amniocentesis. Targeted sequencing of fetal material confirmed that, for all nine couples, our method correctly predicted both the paternally and maternally inherited *CFTR* alleles (Figures 4E and S10). Further, the results also demonstrated that 99.08% of the 2,826 SNPs (including homozygous alleles of class 1 and 2 SNPs) were correctly assigned by TLA to one of the 18 inherited *CFTR* alleles (Table 1).

Next, to demonstrate that this method can readily be adapted to other genomic loci in addition to *CFTR*, we focused our MG-NIPD analysis on *CYP21A2*. Haplotyping the *CYP21A2* locus is more complex due to the nearby homologous pseudogene (*CYP21A1P*), which can induce ambiguous mapping of sequence reads. To circumvent this problem, we made sure that our TLA viewpoint primers uniquely mapped to *CYP21A2* sequences and we sequenced longer (150 and 300 bp) reads. Altering the primers and extending the read length enabled unambiguous mapping to most *CYP21A2* sequences (Figure S11A). We applied *CYP21A2* MG-NIPD to two families (CYP 1 and CYP 3) and could predict, with >99.9% confidence, the two inherited alleles for family 3 (Figures S11B–S11F). The accuracy of prediction and of TLA-generated haplotypes was again confirmed by the sequencing of available amniotic cell DNA. Determining haplotypes in CYP 1 was particularly challenging, as the parents shared one haplotype (99% identical) (Figure S11F). This scenario is analogous to many risk families with identical disease mutations (see for example Figure S3H) and will therefore often be observed in clinical practice. In these cases, class 1 SNPs (i.e., when parents are homozygous for different variants) will be (almost) absent, compromising assessment of the fetal fraction. To circumvent this problem, we included an unrelated locus (here, *CFTR*, though in practice one may use a "neutral" locus not associated with disease) in our analysis, which contributed class 1 and 2 SNPs for estimating the fetal fraction (~26%). In class 2 SNPs, for which the father is heterozygous and the mother homozygous, the shared haplotype will never be observed in paternal-unique reads. However, the non-shared haplotype will be observable as usual. Despite the high fetal fraction, the father's unique haplotype was not observed in cfDNA, making it statistically highly likely (see Material and Methods) that he transmitted his shared allele. For the mother's shared allele this implied that there were only type B SNPs (and no type A). Maternal haplotype 2 was significantly overrepresented in these type B SNPs in cfDNA, which enabled us to predict with >99.9% confidence that the fetus inherited the two nearly identical *CYP21A2* alleles, which was confirmed by sequencing of fetal DNA obtained through amniocentesis (Figure S11F).

Finally, we obtained material from 11 families known to carry β-thalassemia (conferred by mutations in the *HBB* gene) (Figures 5A and S12). We performed MG-NIPD as described above, with the inclusion of a viewpoint near the known disease mutations to firmly embed the mutations in their haplotypes, and multiple additional viewpoints across the *HBB* locus (Table S2). Two families were excluded from analysis. In one case (HBB 3), the parental white blood cells appeared degraded during international transport. In a second (HBB 10), we observed a fetal fraction <1%, which was too low for accurate diagnosis. In the nine remaining cases, we were able to robustly predict the inherited fetal disease status: four fetuses were affected by β-thalassemia, four carried the maternal disease allele, and one fetus was unaffected by β-thalassemia (Figures 5A–5F and S13). All predictions were confirmed by parallel invasive diagnostic tests (Figures 5G and S14), which also showed that 3,639 out of the 3,699 (>98%) neutral and verifiable SNPs were correctly phased by TLA technology to either the mutated or the wild-type allele.

(C) Six class 1 SNPs were identified, with a median of 4.5% paternal alleles, indicating a ~9% fetal fraction.
(D) In 220 SNPs, the paternal WT haplotype is visible, and representing 3.9% of sequence reads, indicating a FF of ~7.8%. In 67 SNPs, the paternal affected allele would be visible if it was inherited, but only 0.1% of reads are of the paternal allele, likely due to sequence errors.
(E) Maternal heterozygous SNPs were classified type A where the maternal WT allele was identical to the paternal inherited allele and type B where the maternal affected allele was identical to the paternal inherited allele. The distribution of the affected allele is shown left of the dotted line, assuming either maternal allele is inherited. The observed distributions of A and B SNPs is shown on the right of the dotted line and clearly corresponds to the expected distribution if the maternal affected allele is inherited. Squares indicate the mean fraction of hap2 alleles, whiskers indicate the 95% confidence interval, corrected for overdispersion.
(F) The probability of each of the four possible combinations of haplotypes is calculated, in this case the probability of a heterozygous fetus carrying maternal affected allele is >99.9%.
(G) The inherited haplotypes were merged to a predicted fetal genotype, spanning 938 SNPs. This genotype corresponds very strongly to the genotype observed after sequencing the CVS sample obtained directly from the fetus, indicating highly accurate haplotyping and a correct genotype prediction.
(H) Probability calculations for the other eight involved β-thalassemia risk families.

Discussion

In this work we have demonstrated that the combination of targeted haplotyping of the two parents with targeted sequencing of cell-free DNA extracted during pregnancy allows for robust non-invasive prenatal diagnosis of monogenic diseases, without the need to include (or even have) a first affected child for further genetic characterization. We expect this will increasingly be recognized as favorable now that pre-conception screening programs for severe Mendelian disorders are being implemented in our health care system, which inform young couples about their carrier status even before the birth of a proband. Recently, another study appeared which also demonstrated that NIPD can be carried out without including a proband for genetic characterization.[21] Different from our strategy that relies on targeted haplotyping of the locus of interest though, their approach relies on 10× genomics-based whole-genome haplotyping of both parents. This requires purchasing the necessary equipment. It also seems to make sequencing unnecessarily expensive and data analysis and storage computationally more demanding. In recognition thereof, the authors proposed that prior to haplotype sequencing, a probe-based capture step can be incorporated to specifically direct sequencing to the locus of interest.[21] The future will tell whether in terms of sensitivity and accuracy, cost effectiveness, and readiness to implement, either of the two approaches is to be preferred over the other.

We believe an advantage of MG-NIPD is that it involves targeted sequencing only of the gene of interest. This excludes the possibility of incidentally finding disease mutations in other genes elsewhere in the genome, which many clinicians will perceive as complicating and therefore undesired during pregnancy. The targeted nature of MG-NIPD also makes sequencing costs limited: haplotyping requires only ~3 million read pairs per parent, with cfDNA sequencing requiring on average 9 million reads per family (Table S4). The robustness of predictions made by our method lies in the fact that it tests not only the likelihood of a given allele being transmitted but also the likelihood of the second allele not being transmitted, with each event contributing similarly to the final risk calculation. Thus, for a false diagnosis to be made, the non-inherited haplotype would have to be scored as being overrepresented *and* the inherited haplotype as not being over-represented, which is highly unlikely if the latter haplotype is contributed by the fetus to the maternal blood. Therefore, it is far more likely that MG-NIPD produces an inconclusive, rather than a false-positive or false-negative, prediction. In fact, in the family carrying *HBB* with an extremely low (1%) fetal cfDNA fraction, an inconclusive analysis was precisely the result. The same may happen when meiotic recombination has rearranged the locus of interest. The chance of this happening inside the relatively small loci that we are considering is low, yet one needs to be aware of this possibility. By centering the locus of interest around the disease mutation (rather than having it at the edge of the phased genomic interval), recombination events will lead to inconclusive rather than false predictions. In situations where MG-NIPD results are inconclusive, the couple may still opt for an invasive test. Further validation on larger numbers of pregnancies is needed to determine whether MG-NIPD provides the high degree of accuracy that is needed to eventually replace current invasive strategies for prenatal diagnosis.

For MG-NIPD, no specialized laboratory equipment is required, making this a method that can be readily implemented in any genetics laboratory. The principle shown here is applicable to any risk locus where the disease-causing sequence variant is known and where allele-discerning SNPs are present within the risk locus. Our current work focused on autosomal-recessive disorders, but in principle the strategy should perform equally on autosomal-dominant disorders, also if the mother were the carrier. Embedding triplet repeat expansions in their haplotypes by TLA is less trivial: extending this approach to disorders caused by such expansions would therefore require further optimization. We anticipate that MG-NIPD will be an attractive means to give couples additional comfort early during pregnancy that the embryo selected by pre-implantation genetic diagnostics (PGD) is indeed not affected, which for technical reasons cannot fully be excluded by PGD.[37,38] Finally, modified versions of the MG-NIPD method presented here may offer a non-invasive and easy way to confirm parenthood, for example following *in vitro* fertilization. Thus, in many cases in future prenatal diagnostics, a simple blood test could give desired comfort during pregnancy and replace more burdensome and risky invasive tests such as CVS and amniocentesis. The scripts used in this publication are publicly available through github (see Web Resources).

Accession Numbers

Sequence data used in this research has been deposited at the European Genome-phenome Archive (EGA), which is hosted by the EBI and the CRG, under accession number EGAS00001002622.

Supplemental Data

Supplemental Data include 14 figures, 4 tables, and a technical appendix and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2017.07.012.

Conflicts of Interest

C. Vermeulen, M.J.A.M.V., and G.G. are shareholders of Cergentis. E.d.W. and E.S. are co-founders and shareholders of Cergentis. W.d.L. is co-founder, shareholder, and scientific advisor of Cergentis and holds a patent application on TLA (WO2012005595).

Acknowledgments

We thank Utrecht Sequencing Facility (USF) for providing sequencing data and service, Carien Hilvering for providing a

custom primer design tool, P.I. de Bakker for helpful discussion and critical reading of the manuscript, Michael van Gerven for designing part of the primers, and Ewart Kuijk for providing cell lines used in optimization experiments. This work was supported by a grant from the U-fonds and the K.F. Hein Fonds, an NWO/CW TOP grant (714.012.002), an NWO VICI grant (724.012.003), and an EU grant 2010-259743 (MODHEP) to W.d.L.

Received: June 16, 2017
Accepted: July 24, 2017
Published: August 24, 2017

Web Resources

European Genome-phenome Archive (EGA), https://www.ebi.ac.uk/ega
GenBank, http://www.ncbi.nlm.nih.gov/genbank/
Globin Gene Server, http://globin.cse.psu.edu/
MG-NIPD, https://github.com/deLaatLab/MG-NIPD
NCBI, http://www.ncbi.nlm.nih.gov/
OMIM, http://www.omim.org/
UCSC Genome Browser, http://genome.ucsc.edu

References

1. Diaz, L.A., Jr., and Bardelli, A. (2014). Liquid biopsies: genotyping circulating tumor DNA. J. Clin. Oncol. *32*, 579–586.
2. Diehl, F., Schmidt, K., Choti, M.A., Romans, K., Goodman, S., Li, M., Thornton, K., Agrawal, N., Sokoll, L., Szabo, S.A., et al. (2008). Circulating mutant DNA to assess tumor dynamics. Nat. Med. *14*, 985–990.
3. Lo, Y.M., Corbetta, N., Chamberlain, P.F., Rai, V., Sargent, I.L., Redman, C.W., and Wainscoat, J.S. (1997). Presence of fetal DNA in maternal plasma and serum. Lancet *350*, 485–487.
4. Lo, Y.M., Tein, M.S., Lau, T.K., Haines, C.J., Leung, T.N., Poon, P.M., Wainscoat, J.S., Johnson, P.J., Chang, A.M., and Hjelm, N.M. (1998). Quantitative analysis of fetal DNA in maternal plasma and serum: implications for noninvasive prenatal diagnosis. Am. J. Hum. Genet. *62*, 768–775.
5. Shi, X., Zhang, Z., Cram, D.S., and Liu, C. (2015). Feasibility of noninvasive prenatal testing for common fetal aneuploidies in an early gestational window. Clin. Chim. Acta *439*, 24–28.
6. Chiu, R.W., Chan, K.C., Gao, Y., Lau, V.Y., Zheng, W., Leung, T.Y., Foo, C.H., Xie, B., Tsui, N.B., Lun, F.M., et al. (2008). Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. Proc. Natl. Acad. Sci. USA *105*, 20458–20463.
7. Fan, H.C., Blumenfeld, Y.J., Chitkara, U., Hudgins, L., and Quake, S.R. (2008). Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. Proc. Natl. Acad. Sci. USA *105*, 16266–16271.
8. Caughey, A.B., Hopkins, L.M., and Norton, M.E. (2006). Chorionic villus sampling compared with amniocentesis and the difference in the rate of pregnancy loss. Obstet. Gynecol. *108*, 612–616.
9. Han, J., Pan, M., Zhen, L., Yang, X., Ou, Y.M., Liao, C., and Li, D.Z. (2014). Chorionic villus sampling for early prenatal diagnosis: Experience at a mainland Chinese hospital. J. Obstet. Gynaecol. *34*, 669–672.
10. Akolekar, R., Beta, J., Picciarelli, G., Ogilvie, C., and D'Antonio, F. (2015). Procedure-related risk of miscarriage following amniocentesis and chorionic villus sampling: a systematic review and meta-analysis. Ultrasound Obstet. Gynecol. *45*, 16–26.
11. Lench, N., Barrett, A., Fielding, S., McKay, F., Hill, M., Jenkins, L., White, H., and Chitty, L.S. (2013). The clinical implementation of non-invasive prenatal diagnosis for single-gene disorders: challenges and progress made. Prenat. Diagn. *33*, 555–562.
12. Ferrari, M., Carrera, P., Lampasona, V., and Galbiati, S. (2015). New trend in non-invasive prenatal diagnosis. Clin. Chim. Acta *451* (Pt A), 9–13.
13. Lo, Y.M., Hjelm, N.M., Fidler, C., Sargent, I.L., Murphy, M.F., Chamberlain, P.F., Poon, P.M., Redman, C.W., and Wainscoat, J.S. (1998). Prenatal diagnosis of fetal RhD status by molecular analysis of maternal plasma. N. Engl. J. Med. *339*, 1734–1738.
14. Finning, K., Martin, P., and Daniels, G. (2004). A clinical service in the UK to predict fetal Rh (Rhesus) D blood group using free fetal DNA in maternal plasma. Ann. N Y Acad. Sci. *1022*, 119–123.
15. Drury, S., Mason, S., McKay, F., Lo, K., Boustred, C., Jenkins, L., and Chitty, L.S. (2016). Implementing non-invasive prenatal diagnosis (NIPD) in a National Health Service Laboratory; from dominant to recessive disorders. Adv. Exp. Med. Biol. *924*, 71–75.
16. Verhoef, T.I., Hill, M., Drury, S., Mason, S., Jenkins, L., Morris, S., and Chitty, L.S. (2016). Non-invasive prenatal diagnosis (NIPD) for single gene disorders: cost analysis of NIPD and invasive testing pathways. Prenat. Diagn. *36*, 636–642.
17. Chen, S., Ge, H., Wang, X., Pan, X., Yao, X., Li, X., Zhang, C., Chen, F., Jiang, F., Li, P., et al. (2013). Haplotype-assisted accurate non-invasive fetal whole genome recovery through maternal plasma sequencing. Genome Med. *5*, 18.
18. Ma, D., Ge, H., Li, X., Jiang, T., Chen, F., Zhang, Y., Hu, P., Chen, S., Zhang, J., Ji, X., et al. (2014). Haplotype-based approach for noninvasive prenatal diagnosis of congenital adrenal hyperplasia by maternal plasma DNA sequencing. Gene *544*, 252–258.
19. Lo, Y.M., Chan, K.C., Sun, H., Chen, E.Z., Jiang, P., Lun, F.M., Zheng, Y.W., Leung, T.Y., Lau, T.K., Cantor, C.R., and Chiu, R.W. (2010). Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. Sci. Transl. Med. *2*, 61ra91.
20. Lam, K.W., Jiang, P., Liao, G.J., Chan, K.C., Leung, T.Y., Chiu, R.W., and Lo, Y.M. (2012). Noninvasive prenatal diagnosis of monogenic diseases by targeted massively parallel sequencing of maternal plasma: application to β-thalassemia. Clin. Chem. *58*, 1467–1475.
21. Hui, W.W., Jiang, P., Tong, Y.K., Lee, W.S., Cheng, Y.K., New, M.I., Kadir, R.A., Chan, K.C., Leung, T.Y., Lo, Y.M., and Chiu, R.W. (2017). Universal haplotype-based noninvasive prenatal testing for single gene diseases. Clin. Chem. *63*, 513–524.
22. Kitzman, J.O., Snyder, M.W., Ventura, M., Lewis, A.P., Qiu, R., Simmons, L.E., Gammill, H.S., Rubens, C.E., Santillan, D.A., Murray, J.C., et al. (2012). Noninvasive whole-genome sequencing of a human fetus. Sci. Transl. Med. *4*, 137ra76.
23. Chan, K.C., Jiang, P., Sun, K., Cheng, Y.K., Tong, Y.K., Cheng, S.H., Wong, A.I., Hudecova, I., Leung, T.Y., Chiu, R.W., and Lo, Y.M. (2016). Second generation noninvasive fetal genome analysis reveals de novo mutations, single-base parental inheritance, and preferred DNA ends. Proc. Natl. Acad. Sci. USA *113*, E8159–E8168.

24. New, M.I., Tong, Y.K., Yuen, T., Jiang, P., Pina, C., Chan, K.C., Khattab, A., Liao, G.J., Yau, M., Kim, S.M., et al. (2014). Noninvasive prenatal diagnosis of congenital adrenal hyperplasia using cell-free fetal DNA in maternal plasma. J. Clin. Endocrinol. Metab. *99*, E1022–E1030.
25. Parks, M., Court, S., Bowns, B., Cleary, S., Clokie, S., Hewitt, J., Williams, D., Cole, T., MacDonald, F., Griffiths, M., et al. (2017). Non-invasive prenatal diagnosis of spinal muscular atrophy by relative haplotype dosage. European journal of human genetics. Eur. J. Hum. Genet. *25*, 416–422.
26. de Vree, P.J., de Wit, E., Yilmaz, M., van de Heijning, M., Klous, P., Verstegen, M.J., Wan, Y., Teunissen, H., Krijger, P.H., Geeven, G., et al. (2014). Targeted sequencing by proximity ligation for comprehensive variant detection and local haplotyping. Nat. Biotechnol. *32*, 1019–1025.
27. Vogelstein, B., and Kinzler, K.W. (1999). Digital PCR. Proc. Natl. Acad. Sci. USA *96*, 9236–9241.
28. Regan, J.F., Kamitaki, N., Legler, T., Cooper, S., Klitgord, N., Karlin-Neumann, G., Wong, C., Hodges, S., Koehler, R., Tzonev, S., and McCarroll, S.A. (2015). A rapid molecular approach for chromosomal phasing. PLoS ONE *10*, e0118270.
29. Lun, F.M., Tsui, N.B., Chan, K.C., Leung, T.Y., Lau, T.K., Charoenkwan, P., Chow, K.C., Lo, W.Y., Wanapirak, C., Sanguansermsri, T., et al. (2008). Noninvasive prenatal diagnosis of monogenic diseases by digital size selection and relative mutation dosage on DNA in maternal plasma. Proc. Natl. Acad. Sci. USA *105*, 19920–19925.
30. Cutting, G.R. (2015). Cystic fibrosis genetics: from molecular understanding to clinical application. Nat. Rev. Genet. *16*, 45–56.
31. Speiser, P.W., and White, P.C. (2003). Congenital adrenal hyperplasia. N. Engl. J. Med. *349*, 776–788.
32. Dekkers, J.F., Wiegerinck, C.L., de Jonge, H.R., Bronsveld, I., Janssens, H.M., de Winter-de Groot, K.M., Brandsma, A.M., de Jong, N.W., Bijvelds, M.J., Scholte, B.J., et al. (2013). A functional CFTR assay using primary cystic fibrosis intestinal organoids. Nat. Med. *19*, 939–945.
33. Genome of the Netherlands Consortium (2014). Whole-genome sequence variation, population structure and demographic history of the Dutch population. Nat. Genet. *46*, 818–825.
34. Milani, R., Marcellini, A., Montagner, G., Baldisserotto, A., Manfredini, S., Gambari, R., and Lampronti, I. (2015). Phloridzin derivatives inhibiting pro-inflammatory cytokine expression in human cystic fibrosis IB3-1 cells. Eur. J. Pharm. Sci. *78*, 225–233.
35. Morral, N., Bertranpetit, J., Estivill, X., Nunes, V., Casals, T., Giménez, J., Reis, A., Varon-Mateeva, R., Macek, M., Jr., Kalaydjieva, L., et al. (1994). The origin of the major cystic fibrosis mutation (delta F508) in European populations. Nat. Genet. *7*, 169–175.
36. Liao, G.J., Lun, F.M., Zheng, Y.W., Chan, K.C., Leung, T.Y., Lau, T.K., Chiu, R.W., and Lo, Y.M. (2011). Targeted massively parallel sequencing of maternal plasma DNA permits efficient and unbiased detection of fetal alleles. Clin. Chem. *57*, 92–101.
37. Traeger-Synodinos, J. (2017). Pre-implantation genetic diagnosis. Best Pract. Res. Clin. Obstet. Gynaecol. *39*, 74–88.
38. Wilton, L., Thornhill, A., Traeger-Synodinos, J., Sermon, K.D., and Harper, J.C. (2009). The causes of misdiagnosis and adverse outcomes in PGD. Hum. Reprod. *24*, 1221–1228.